

소량의 분해된 RNA를 이용한 cDNA library 제작

# SMARTer<sup>®</sup> Universal Low Input RNA Kit for Sequencing

- 분해된 RNA(RIN 2-3), non-polyadenylated RNA를 이용하여 cDNA library 제작
- 소량의 시료 - 200 pg rRNA-depleted RNA(2 ng total RNA)
- FFPE나 LCM과 같은 어려운 샘플에 이상적
- 5' 또는 3' 편향이 없는 complete transcriptome coverage
- 뛰어난 재현성, mappability, ERCC correlation

**서론**

차세대 시퀀싱(Next Generation Sequencing, NGS)은 전체 transcriptome에서 높은 감도와 다양한 범위로 RNA 발현 분석을 가능하게 했다. 하지만 FFPE(formaldehyde fixed paraffin embedded tissue)나 LCM(laser captured microdissection)과 같이 RNA 샘플이 분해되어 품질이 좋지 않거나 소량인 샘플에서는 적용하기 어려운 한계가 있다.

SMARTer Universal Low Input RNA Kit for Sequencing을 이용하면 분해되거나 소량의 샘플에서도 random primer를 이용하여 cDNA library를 제작할 수 있고, 이때 이용되는 샘플 양은 불과 2 ng total RNA이다. 이 키트의 산물은 5' 또는 3'으로 편향되지 않은 whole transcriptome coverage를 보이며, 뛰어난 재현성(reproducibility, mappability, ERCC correlation)을 보인다. 또한 transcriptome profiling용 cDNA library 제작은 2일 밖에 소요되지 않는다(그림 1). 또한 본 키트를 이용하여 RIN(RNA Integrity number) value가 2.4인 human brain total RNA 샘플로 cDNA library를 제작하는 실험도 진행되었다.

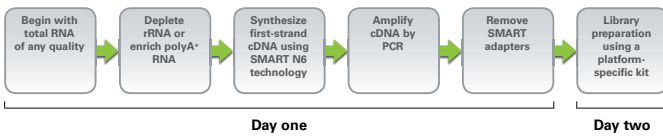


그림 1. NGS Work flow for the SMARTer Universal Low Input RNA Kit for Sequencing

**SMART cDNA 합성 프로토콜**

Clontech의 SMART(Switching Mechanism At the 5' end of the RNA Template) 기술은 단일 반응으로 full-length cDNA를 합성할 수 있는 획기적인 방법이다. 이 방법은 역전사효소의 말단전이효소(terminal transferase) 활성과 주형전환(template switching) 활성을 이용하여 1<sup>st</sup> strand cDNA의 3' 말단에 PCR adaptor를 부가하므로 별도의 ligation 단계가

필요 없다.

잘 보존된 mRNA 샘플의 경우에는 역전사 단계에서 oligo dT primer를 이용하고 single cell 레벨에도 적용 가능한 SMARTer Ultra Low Input RNA Kit for Illumina Sequencing-HV를 추천한다. 반면, 분해된 소량의 RNA(200 pg- 10 ng)를 이용하여 cDNA를 합성하는 실험에는 random primer(modified N6 primer)를 이용하는 SMARTer Universal Low Input RNA Kit for Sequencing을 추천한다(그림 2).

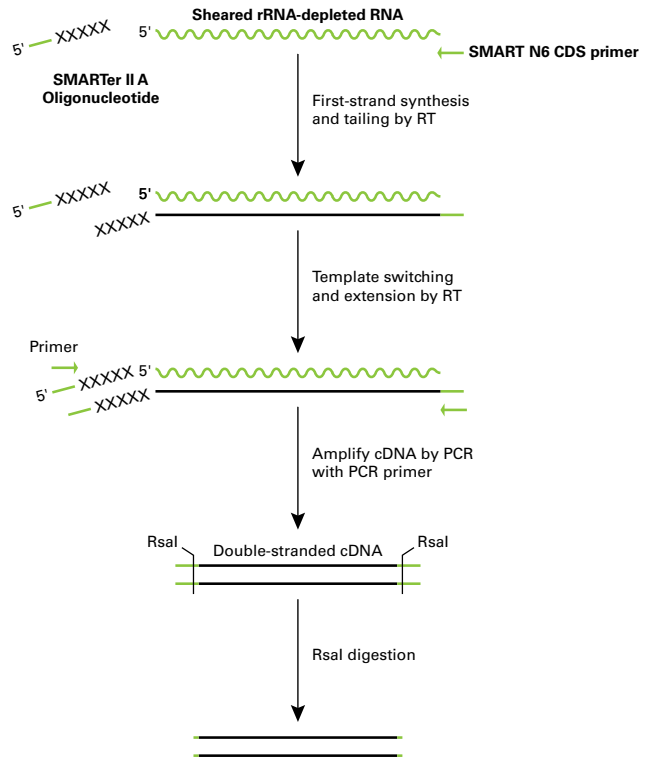
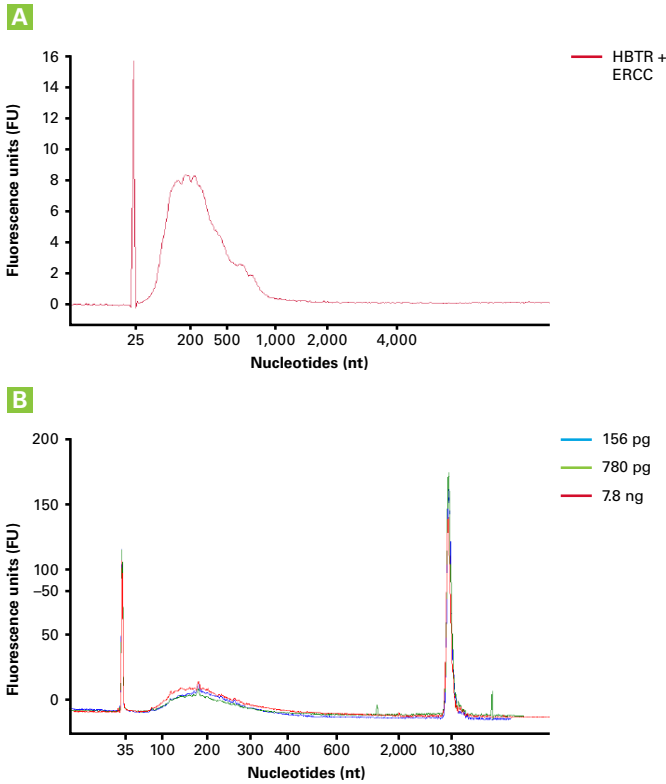


그림 2. Synthesis scheme for SMARTer Universal Low Input RNA Kit for Sequencing. First-strand cDNA is synthesized from sheared rRNA depleted RNA, using a modified N6 primer called SMART N6 CDS (N = A, G, T, or C). When SMARTScribe Reverse Transcriptase reaches the 5' end of the RNA, its terminal transferase activity adds a few additional nucleotides to the 3' end of the cDNA. This non-template nucleotide stretch anneals with the SMARTer Oligonucleotide, creating an extended template where the RT continues through the II A sequence. The SMARTer anchor sequence and the modified N6 sequence serve as universal priming sites for cDNA amplification. Finally, amplified cDNA is digested with RsaI to remove the adapter prior to sequencing.

**분해된 소량의 human brain total RNA을 이용하여 cDNA library 제작**

Human Brain Total RNA(Code 636530)를 화학적으로 절단<sup>(1)</sup>한 후 SMARTer Universal Low Input RNA Kit for Sequencing을 이용하여 cDNA library를 제작하였다. 300 ng의 절단된 human brain RNA에 화학적으로 절단한 ERCC(External RNA Controls Consortium) control RNA를 섞었다. 그 후, 100 ng의 샘플을 Epicentre Ribo-Zero(Illumina, Cat. No. MRZH116)의 변형 프로토콜을 이용하여 rRNA를 제거했다. 분해된 RNA는 electropherogram에서 200 bp 근처에서 넓은 peak를 나타내었다(그림 3, panel A).

**Note:** Epicentre Ribo-Zero의 변형 프로토콜은 Clontech([www.clontech.com](http://www.clontech.com))에서 "Protocol for Removal of rRNA from Small Amounts of Total RNA"로 검색하면 자세한 내용을 확인할 수 있다.



**그림 3. Yield and purity of input RNA and resulting cDNA libraries.** Panel A. Human Brain Total RNA (HBTR) was chemically sheared, spiked with ERCC control RNA (4 µl of a 1:1,000 dilution per 100 ng), and rRNA-depleted using a modified Ribo-Zero protocol for lowinput samples. One microliter was analyzed using the Agilent 2100 Bioanalyzer (RNA 6000 Pico chip). Panel B. Three amounts of rRNAdepleted RNA (156 pg, 780 pg, and 7.8 ng) were used as input for the SMARTer Universal Low Input RNA Kit for Sequencing, and 1 µl of the resulting cDNA library was analyzed on the Agilent 2100 Bioanalyzer (High Sensitivity DNA chip).

위와 같이 rRNA를 제거한 RNA를 cDNA 합성에 사용하였다(156 pg, 780 pg, 7.8 ng의 rRNA-cleared RNA는 각각 total RNA 2 ng, 10 ng, 100 ng에 해당하는 양이다). SMARTer Universal Low Input RNA Kit for Sequencing을 이용하여 제작된 cDNA library는 사용한 RNA 농도에 상관없이 비슷한 수율과 순도를 나타냈다(그림 3, panel B). 각

각의 cDNA library 1-2 ng으로 Clontech의 Low Input Library Prep Kit(Code 634947)을 이용하여 Illumina MiSeq 장비로 분석하기 위한 시퀀싱용 library를 준비하였다. 모든 library는 높은 수치로 read되었으며 84-87%의 mapped, unique mapped 77-79%이며, 15,000개 이상의 유전자를 식별했다. 3 종류의 library에서 rRNA는 2% 미만이었다 (표 1).

| Table 1: Sequence Alignment Metrics |                  |                  |                  |
|-------------------------------------|------------------|------------------|------------------|
| Input RNA                           | 156 pg           | 780 pg           | 7.8 ng           |
| No. of Reads                        | 27,043,029       | 25,247,363       | 16,991,089       |
| Mapped to rRNA                      | 316,377 (1.2%)   | 360,696 (1.4%)   | 324,628 (1.9%)   |
| Mapped to Mitochondrial RNA         | 6,408,693 (24%)  | 6,007,327 (24%)  | 3,858,048 (23%)  |
| Mapped to RefSeq                    | 16,863,730 (84%) | 16,158,298 (86%) | 10,977,858 (87%) |
| Uniquely Mapped to RefSeq           | 15,480,061 (77%) | 14,823,088 (79%) | 10,034,740 (79%) |
| Exons                               | 6,464,355 (38%)  | 6,185,920 (38%)  | 4,256,916 (39%)  |
| Introns                             | 10,399,375 (62%) | 9,972,378 (62%)  | 6,720,942 (61%)  |
| Genes Identified                    | 15,140           | 15,574           | 15,697           |

**표 1. Sequencing alignment metrics for three N6 libraries.** cDNA libraries were created using 156 pg, 780 pg, and 7.8 ng of rRNA-depleted human brain reference RNA and used for sequencing. Illumina adapters and indices were added to 1-2 ng of each cDNA library using the Low Input Library Prep Kit, and the cDNA samples were sequenced on an Illumina MiSeq Platform with 1 x 50 bp reads. Reads were trimmed by CLC Genomics Workbench and mapped to rRNA and the mitochondrial genome with CLC (% reads indicated). The unmapped reads were subsequently mapped with CLC to the human genome with the RefSeq masking, producing mapped reads and uniquely mapped reads. The number of genes identified in each library was determined by the number of genes with an RPKM (read per kilobase of exon per million of reads) of at least 0.1. The number of reads that map to introns or exons is a percentage of the reads successfully mapped to RefSeq. Note: The 156 pg, 780 pg, and 7.8 ng of rRNAdepleted RNA correspond to 2 ng, 10 ng, and 100 ng of total RNA, prior to rRNA depletion.

SMARTer Universal Low Input RNA Kit for Sequencing은 ~87%의 높은 mappability를 나타내며, 다양한 농도의 RNA 적용에서도 높은 재현성과 민감도를 나타낸다(그림 4). 156 pg-7.8 ng의 rRNA-depleted RNA로 각각 제작한 cDNA library에서 발현 정도에 대한 Pearson correlation data는 높은 재현성을 보여준다(그림 4, Panel A-C). 156 pg의 데이터에서 편차가 약간 증가하는 것은 소량의 RNA를 이용하였기 때문일 것이다. 각각 다른 농도의 RNA를 이용한 결과값을 correlation plot에서 비교했을 때 모든 데이터 세트에서 일관된 결과를 보였다(그림 4, Panel D-F)

이와 같은 결과는 SMARTer Universal Low Input RNA Kit for Sequencing을 이용하면 156 pg만큼의 매우 적은 양의 rRNA-cleared RNA로도 고품질의 cDNA library를 제작할 수 있다는 것을 보여준다.

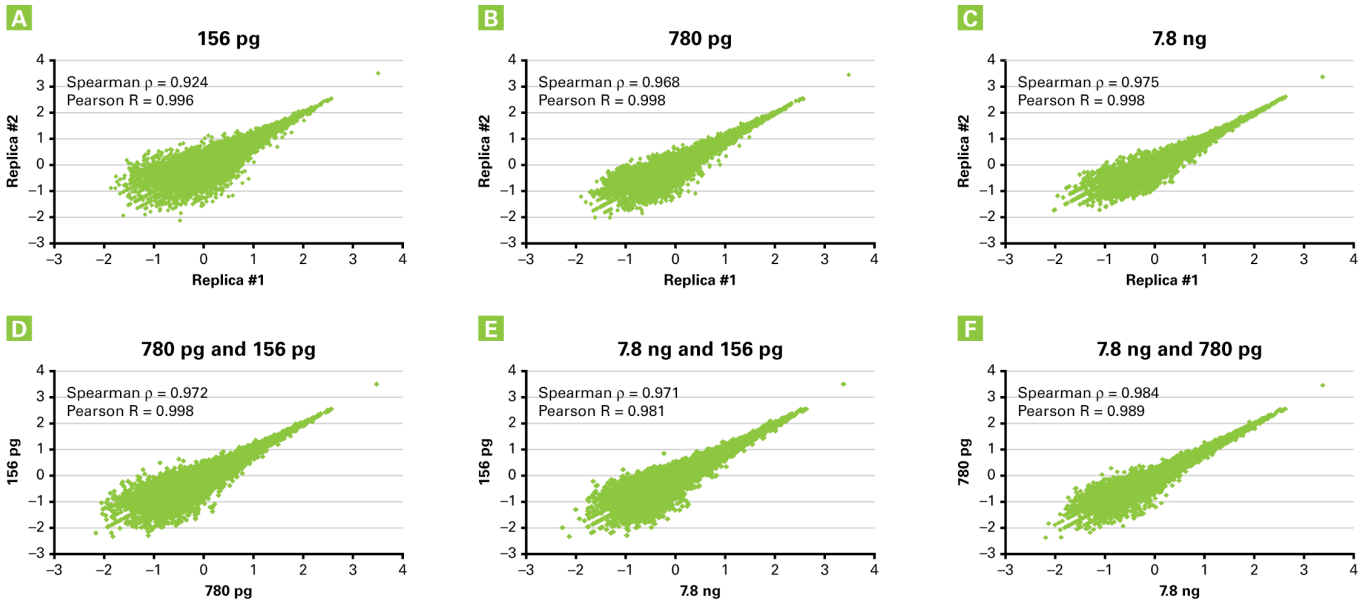


그림 4. High reproducibility and sensitivity across a wide RNA input range. Scatter plots of expression (RPKM, read per kilobase of exon per million of reads) for cDNA libraries prepared from rRNA-depleted human brain RNA. Panels A-C. Comparisons of pairs of cDNA library replicas (Replica #1 and Replica #2) created from 156 pg, 780 pg, and 7.8 ng of input RNA show high reproducibility across a wide range of RNA concentrations. Panels D-F. Comparisons of cDNA libraries generated from pairs of RNA input amounts (780 pg and 156 pg, 7.8 ng and 156 pg, and 7.8 ng and 780 pg) show a high correlation, suggesting consistency across input levels. Axes are plotted on a log<sub>10</sub> scale. Insets indicate the coefficient of correlation by Spearman analysis ( $\rho$ ) and by Pearson correlation ( $R$ ).

ERCC<sup>(2)</sup> 분석에서는 >0.99 Pearson correlation을 나타냄으로써 duplicate library 사이의 높은 재현성을 확인할 수 있었다(그림 5). 또한 시퀀싱을 통해서 밝혀진 발현 레벨은 첨가된 control transcript의 레벨과 일치하는 결과를 보여주었다. 이는 0.84-0.97의 slope와 0.858-0.936의 R<sup>2</sup> 값으로 보여진다. 예상대로 일반적으로 사용하는 RNA 양보다 훨씬 적은 양이 첨가된 control transcript는 나타나지 않았다.

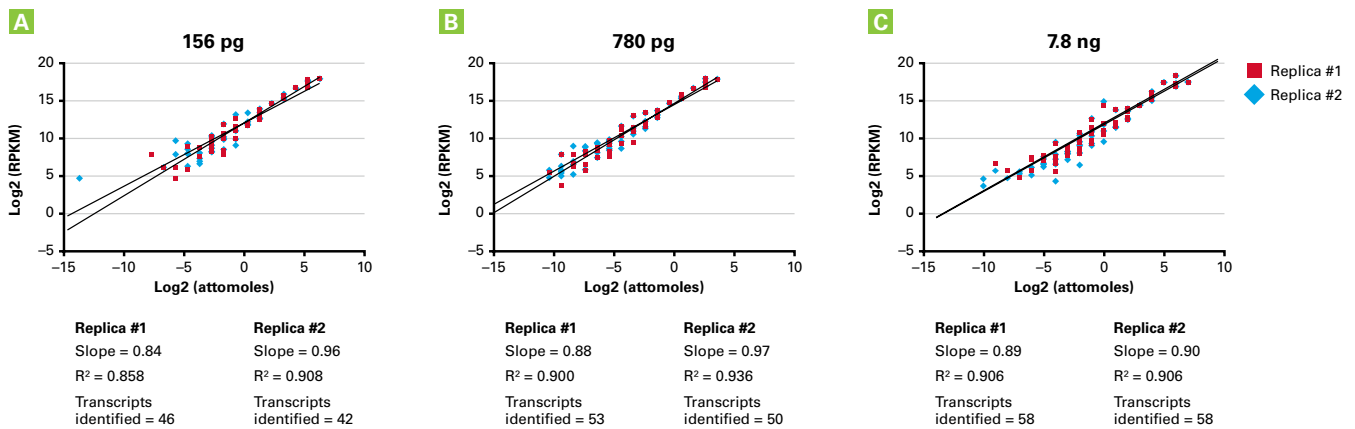
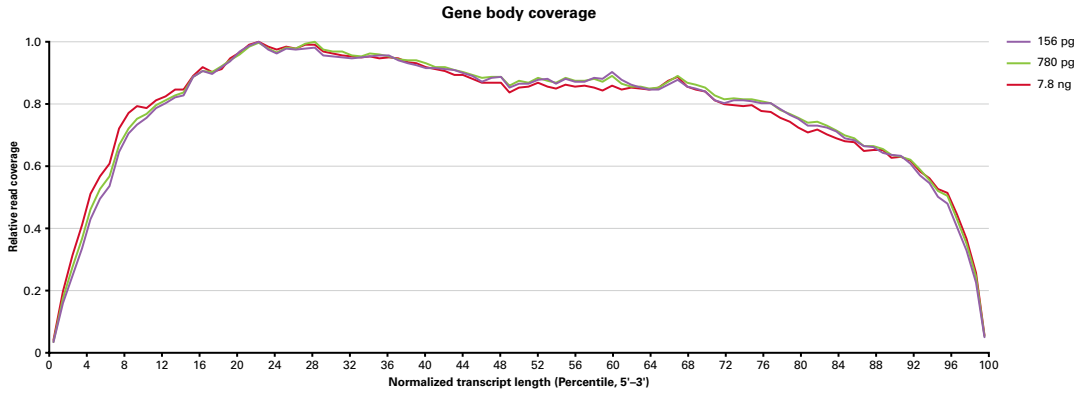


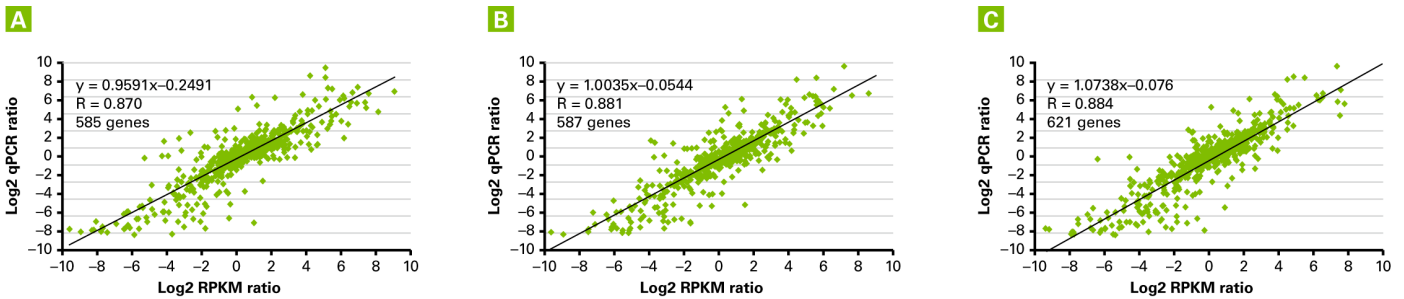
그림 5. High reproducibility confirmed by ERCC analysis. The ERCC (External RNA Control Consortium) Spike-In Mix (92 transcripts) was chemically sheared and 4 microliters (1:1,000 dilution) was spiked into 100 ng of chemically sheared human brain total RNA prior to rRNA depletion. Two cDNA library replicates were generated with each of 156 pg, 780 pg, and 7.8 ng of rRNA-depleted RNA/ERCC mix using the SMARTer Universal Low Input RNA Kit for Sequencing. RPKM of the ERCC transcripts was plotted against the attomoles per transcript spiked in. The slope, R<sup>2</sup>, and the number of transcripts are indicated per replicate. Pearson correlation ( $R$ ) values per replica set were determined to be 0.998, 0.999, and 0.992 for Panels A-C, respectively. Axes are plotted on a log<sub>2</sub> scale.

SMARTer Universal Low Input RNA Kit for Sequencing은 소량의 total RNA로 실험을 시작하더라도 넓은 범위에서 균일한 transcript coverage를 제공한다. 156 pg, 780 pg, 또는 7.8 ng의 human brain total RNA를 이용하여 cDNA library를 제작한 뒤 유전자의 body coverage를 비교하기 위해서 데이터 plot을 겹쳤을 때, 모든 조건에서 5' 또는 3' 편향은 아주 적었다(그림 6).



**그림 6. Uniform gene body coverage.** Overlaid data comparing transcript coverage from cDNA libraries generated using 156 pg, 780 pg, and 7.8 ng of rRNA-depleted human brain RNA. The x-axis represents gene length (RefSeq) normalized to 100%, where 0 is the 5'-end of each transcript and 100 is the 3'-end. The y-axis represents the read coverage relative to the highest coverage percentile.

매우 적은 양의 RNA를 SMARTer Universal Low Input RNA Kit for Sequencing으로 합성한 cDNA는 훨씬 많은 양의 RNA를 필요로 하는 전통적인 방법(e.g. microarray)과 유사한 gene count를 나타낸다. SMARTer Universal Low Input RNA Kit for Sequencing으로 얻어진 데이터와 MicroArray Quality Control(MAQC) project<sup>(3)</sup>을 이용하여 얻어진 qPCR 데이터는 높은 상관성을 보여준다(그림 7). 159 pg에서 5.3 ng의 RNA 범위를 이용하여 얻은 결과값이 높은 상관성(R 0.87-0.88)을 나타내며, 이는 본 키트를 사용하여 얻은 결과값이 매우 정확하다는 것을 보여준다고 할 수 있다.



**그림 7. High correlation with qPCR data via MAQC analysis.** Scatter plots were used to compare differential expression data obtained by sequencing cDNA libraries created with the SMARTer Universal RNA Kit [106 pg HURR/159 pg HBRR (A), 530 pg HURR/790 pg HBRR (B), and 5.3 ng HURR/7.9 ng HBRR (C)] and quantitative PCR (qPCR) data available for HURR and HBRR through the MAQC (MicroArray Quality Control) project (2). The slope and correlation (R) for the comparison line of expression ratio (in RPKM) and qPCR ratio (in Ct) are plotted for HURR and HBRR, in a log<sub>2</sub> scale. The number of genes present in both data sets is indicated. HURR = Human Universal Reference RNA, HBRR= Human Brain Reference RNA.

**결론**

SMARTer Universal Low Input RNA Kit for Sequencing은 매우 적은 양의 분해된 total RNA(RIN value 2-3)나 poly A가 없는 mRNA를 이용하여 효과적으로 RNA-seq를 위한 library를 제작할 수 있다. 이 기술은 최소 200 pg의 rRNA-cleared RNA(2 ng total RNA)만으로도 RNA-Seq에 적용할 수 있으며, 따라서 FFPE나 LCM 분석에 이상적이다. SMARTer 기술은 유전자 coverage에서 최소의 편향을 나타내며, 기존의 방법보다 뛰어난 재현성, mappability, ERCC correlation을 보인다.

**Reference**

- (1) Mortazavi, A. et al.(2010) *Nat. Methods* 5(7):621-628.
- (2) Jiang, L. et al. (2011) *Genome Res.* 21(9):1543-1551.
- (3) MAQC Consortium (2006) *Nat. Biotechnol.* 24(98):1151-1161.

**관련제품**

| Code   | 제품명  | 용량       |
|--------|--|----------|
| 634938 | SMARTer Universal Low Input RNA Kit for Sequencing       | 10 Rxns  |
| 634940 |  | 25 Rxns  |
| 634945 | SMARTer Universal Low Input RNA Library Prep Kit         | 10 Rxns  |
| 634946 |  | 25 Rxns  |
| 634820 | SMARTer Ultra Low Input RNA for Illumina Sequencing – HV | 12 Rxns  |
| 634823 |  | 24 Rxns  |
| 634826 |  | 48 Rxns  |
| 634828 |  | 96 Rxns  |
| 634830 |  | 480 Rxns |
| 634936 | SMARTer Ultra Low RNA Kit for Illumina Sequencing        | 100 Rxns |
| 634947 | Low Input Library Prep Kit                               | 12 Rxns  |
| 634832 | SMARTer Ultra Low RNA Kit for the Fluidigm C1 System     | 2 IFCs   |
| 634833 |  | 10 IFCs  |